

VARIABLES SINTOMÁTICAS EN LAS ESTIMACIONES POBLACIONALES A NIVEL CANTONAL EN COSTA RICA

Edwin A. Chaves Esquivel

Universidad de Costa Rica

Universidad Nacional Autónoma de Costa Rica

RESUMEN

Dado que Costa Rica no realizó el censo de población y vivienda correspondiente a la ronda de 1990, ha sido necesario encontrar técnicas que permitan estimar sus principales características demográficas. En este artículo se analiza el uso de variables tales como el padrón electoral, los datos vitales y la matrícula escolar en la realización de estimaciones poblacionales a nivel cantonal. Para ello se aplicaron cuatro métodos que demostraron ser buenas alternativas.

**SYMPTOMATIC VARIABLES IN POPULATION
ESTIMATES AT THE CANTON LEVEL
IN COSTA RICA**

ABSTRACT

Since Costa Rica did not carry out a population and housing census for the 1990 round, it has been necessary to find other methods for arriving at estimates of the country's main demographic features. This article analyses the use of such variables as voting lists, vital statistics and school enrolment records in the preparation of population estimates at the canton level. To this end, four methodologies that proved to be satisfactory options were applied.

INTRODUCCIÓN

Debido a la problemática que conllevan las estimaciones poblacionales a nivel de áreas administrativas menores, desde hace muchas décadas diversos investigadores han propuesto utilizar variables sintomáticas. Éstas son variables auxiliares ligadas al crecimiento poblacional, que pueden ser utilizadas para medir sus cambios, y que se construyen a partir de información recolectada regularmente por empresas e instituciones públicas o privadas con fines de orden administrativo; se trata, entre otras, de la matrícula escolar, las declaraciones de impuestos, el número de abonados a los servicios básicos y de electores inscritos en el padrón electoral (Serow y Rives, 1995; Simpson y otros, 1996).

La intención de recolectar estos datos no es estimar poblaciones. Esto obliga a estar pendiente de ciertos problemas como el cambio de las definiciones en el tiempo y las interrupciones en la recopilación, que pueden restringir las posibilidades de aplicación con fines estimativos.

Las técnicas utilizadas para trabajar con estas variables son muy variadas. Evidentemente, la precisión de las estimaciones depende de la relación existente entre las variables sintomáticas y el tamaño de la población, así como del grado de relación a lo largo del tiempo, pero también influye la disponibilidad de datos con el detalle debido.

El presente estudio ofrece una contribución en este sentido. Se determinan las variables sintomáticas más estrechamente asociadas al crecimiento poblacional, y se aplican cuatro métodos de estimación que han dado resultados aceptables en países desarrollados. Asimismo, se efectúan estimaciones poblacionales a nivel cantonal para 1984 (año del último censo), que luego se comparan con su valor censal, para medir la precisión de cada uno de los indicadores.

I. ESTRATEGIA METODOLÓGICA

El presente estudio se basa en la aplicación de cuatro métodos que, por las características del medio y de la información disponible, pueden ofrecer aportes importantes en esta materia, a nivel nacional. Estos métodos permiten hacer estimaciones para las que se requiere conocer el tamaño de la población de un área mayor que comprenda una serie de áreas más pequeñas. Por tal razón, las estimaciones deben prorratearse, de manera que el conjunto de ellas reproduzca la población total del área.

1. Método de razón censal

En este método se consideran las tasas de ocurrencia de un determinado indicador sintomático, partiendo del supuesto de que las áreas locales presentan una razón de cambio proporcional a la del área mayor, en el período comprendido entre el último censo y la fecha de estimación (Voss y otros, 1995).

En primer lugar, el procedimiento consiste en determinar la tasa de ocurrencia de cada variable sintomática para las distintas áreas menores, durante el último período censal. Si se desea estimar la población a partir de la variable sintomática S para un cantón u , t años después del último censo, se utiliza la siguiente ecuación:

$$r_0(u) = \frac{S_0(u)}{P_0(u)} \quad [1]$$

$r_0(u)$ es la tasa de ocurrencia del indicador sintomático S para u en el año del último censo

$S_0(u)$ es la información de la variable sintomática S observado para u en el año del último censo

$P_0(u)$ es la población del cantón u medida en el último censo.

Posteriormente, se aproxima la tasa de la variable sintomática a la fecha t , por medio de un parámetro ϕ , que representa la razón de cambio en la variable sintomática entre el año censal y la fecha t , de tal modo que:

$$r_t(u) = \phi r_0(u) \quad [2]$$

Para estimar el valor de ϕ es necesario tener información del área mayor que abarque a u , sobre la cual se pueden hacer estimaciones independientes de la población en la fecha t . Esta área se designa como M y su población en la fecha t como $P_t(M)$, donde la estimación de la población de M en t es $\hat{P}_t(M)$. De este modo, se calculan las tasas de ocurrencia de S en el área M , en la fecha del censo y la fecha t , utilizando la fórmula [1]. Si la razón de cambio en la variable sintomática para el área mayor M puede aproximarse efectivamente al valor de ϕ , se tiene:

$$\hat{\phi} = \frac{\hat{r}_t(M)}{r_0(M)} \quad [3]$$

Entonces, la estimación de $r_t(u)$ es:

$$\hat{r}_t(u) = \hat{\phi} r_0(u) = \frac{\hat{r}_t(M)}{r_0(M)} r_0(u) \quad [4]$$

Por lo tanto, la estimación poblacional para cada cantón u , en t , viene dada por:

$$\hat{P}_t(u) = \frac{S_t(u)}{\hat{r}_t(u)} \quad [5]$$

Generalmente, se obtiene más de una estimación de $P_t(u)$ al considerar diferentes variables sintomáticas. La estimación final es, por lo general, el promedio simple de las estimaciones individuales, aunque hay métodos para ponderarlas (Ericksen, 1973; Ericksen, 1974).

La principal ventaja de este método es la simplicidad de su aplicación; su desventaja es que la dirección de las series sintomáticas podría verse afectada por factores distintos a los que justifican el cambio en la población. Esto significa que la razón para un cantón puede cambiar en una proporción diferente y podría, incluso,

evolucionar en dirección contraria a la razón del área mayor; en tal caso, no se cumpliría el supuesto básico del modelo.

2. Método de diferencia de tasas

Este procedimiento es una variación del método de razón censal. La principal diferencia entre los dos métodos radica en la forma de estimar la tasa de ocurrencia de la variable sintomática para cada cantón. De acuerdo a este método se supone que el crecimiento de esta tasa para el cantón u es igual al crecimiento registrado en el área mayor M .

Considerando nuevamente el cálculo de estimaciones por medio de la variable sintomática S , la aproximación para la tasa de ocurrencia de este indicador en la fecha t , para el cantón u , viene dada por la siguiente fórmula:

$$\hat{r}_t(u) = r_0(u) + [\hat{r}_t(M) - r_0(M)] \quad [6]$$

donde sus componentes se definen como se especificó anteriormente.

3. Método de correlación de razón

La aplicación de este método consiste fundamentalmente en dividir la población de un área mayor en los cantones que la constituyen. Esta técnica ha demostrado ser una de las más exactas en la estimación de población poscensal (Namboodiri, 1972; Mandell y Tayman, 1982; Fenney, Hibbs y Gillaspay, 1995).

El método se basa en el supuesto de que existe una relación lineal entre las razones de cambio de las proporciones poblacionales y las razones de cambio de las proporciones de las variables sintomáticas entre dos períodos. Considerando, además, que los cambios en esta relación a lo largo del tiempo afectan uniformemente todos los cantones, se desarrolla un modelo más refinado, que puede formularse de la siguiente manera:

$$Y_u = a_0 + a_1 X_u \quad [7]$$

$$\text{donde: } Y_u = \frac{\frac{P_1(u)}{P_1(M)}}{\frac{P_0(u)}{P_0(M)}} \quad \text{y} \quad X_u = \frac{\frac{S_1(u)}{S_1(M)}}{\frac{S_0(u)}{S_0(M)}} \quad \text{para todo cantón } u$$

a_i ($i = 0, 1$) corresponde a los coeficientes del modelo.

Los valores P , S , u y M están definidos tal como se indicó en los métodos anteriores y sus subíndices 0 y 1 representan las fechas en que se realizaron el penúltimo y el último censo, respectivamente.

Esta ecuación indica que la razón de cambio a lo largo del tiempo de la proporción poblacional está sistemáticamente relacionada con la correspondiente razón de cambio en las proporciones de los indicadores. Se pueden introducir en el modelo otras variables sintomáticas, de tal manera que:

$$Y_u = a_0 + \sum_{j=1}^m a_j X_{u,j} \quad \text{donde} \quad X_{u,j} = \frac{\frac{S_{j,1}(u)}{S_{j,1}(M)}}{\frac{S_{j,0}(u)}{S_{j,0}(M)}} \quad [8]$$

para cada cantón u y cada variable sintomática j .

Una vez estimado el modelo de regresión, es muy sencillo hacer las estimaciones de población para un período poscensal.

4. Método de correlación de tasa

Se le critica a la técnica de correlación de razón una inconsistencia interna, ya que los coeficientes del modelo se calculan tomando como referencia el tiempo transcurrido entre los censos. Sin embargo, las estimaciones no coinciden con el período base, que por lo general, es menor que el tiempo transcurrido entre los censos. Esto puede afectar la estabilidad natural de los coeficientes y las estimaciones producidas (Mandell y Tayman, 1982; O'Hare, 1980; Swanson, 1980).

Se ha propuesto una modificación del método de correlación de razón llamada “tasa de correlación”, que se basa en una aproximación exponencial de cambio y se realiza tomando el logaritmo natural de cada razón y dividiéndolo por el tiempo transcurrido entre los censos utilizados como referencia (Swanson y Tedrow, 1984). Formalmente, el modelo se puede expresar de la siguiente manera:

$$Y_u^* = a_0 + \sum_{j=1}^m a_j X_{u,j}^* \quad [9]$$

donde: $Y_u^* = \frac{\ln(Y_u)}{k}$ y $X_{u,j}^* = \frac{\ln(X_{u,j})}{k}$

Y_u y $X_{u,j}$ están definidos tal como se indicó en el caso de la técnica de correlación de razón y k es el intervalo de tiempo transcurrido entre los censos.

Después de un experimento que se llevó a cabo en varios condados del estado de Washington, Swanson y Katzoff señalan tres razones para suponer que la técnica de correlación de tasa da resultados más precisos, que la de correlación de razón:

i) Como la primera proviene de una transformación de la segunda, se deduce que ambas tienen la misma tendencia. Si ésta se determina durante varios años de estimaciones, el método de correlación de razón presentará mayores inconsistencias, porque estas extrapolaciones están basadas en una relación estructural de cambio a lo largo de un período mayor de años.

ii) La transformación de las tasas reduce las variancias, por lo que se podría lograr una mayor exactitud con la correlación de tasa.

iii) Puesto que se ha reducido el efecto de la autocorrelación espacial, es de esperar que las estimaciones sean más precisas (Swanson y Tedrow, 1984).

5. Variables sintomáticas disponibles

Para cumplir con los objetivos del estudio es requisito fundamental contar con información sintomática sobre las diferentes áreas

administrativas del país. Debido a las características de los métodos de estimación empleados, era necesario que los datos utilizados estuvieran disponibles a partir del censo de población de 1963. Se seleccionaron siete fuentes de datos: matrícula escolar en primer y segundo ciclos, proporcionada por el Ministerio de Educación Pública; nacimientos; defunciones; permisos de construcción de viviendas, suministrados por la Dirección General de Estadística y Censos; registros de asegurados de la Caja Costarricense del Seguro Social; registros de atención en los centros de salud facilitados por el Ministerio de Salud y número de electores inscritos en el padrón electoral, proporcionado por el Tribunal Supremo de Elecciones.

Cada una de estas fuentes de datos debió ser sometida a un proceso de revisión y evaluación, con el fin de evitar inconsistencias (Chaves, 1997).

6. Valoración de las estimaciones

La mejor manera de evaluar la precisión de las estimaciones producidas por los diferentes métodos, es compararlas con sus valores reales. Sin embargo, debido a que este valor es desconocido, la comparación se hace con los valores censales. Para ello, fue necesario conseguir estimaciones poblacionales correspondientes a 1984, año del último censo poblacional. Éstas se realizaron aplicando los diferentes métodos de estimación, con todas las variables sintomáticas seleccionadas y combinaciones de éstas.

Para medir la calidad de las estimaciones producidas se calculó el error porcentual (*EP*) o porcentaje de desviación respecto del valor censal. La medición se realiza aplicando la siguiente fórmula:

$$EP = \frac{(\text{Población estimada} - \text{Población censal})}{\text{Población censal}} \times 100 \quad [10]$$

Un error porcentual positivo indica que en la estimación se sobreestima la población censal, mientras que un valor negativo indica una subestimación.

Puesto que las estimaciones se prorratan para ajustar la población total del país, el error porcentual promedio presenta valores cercanos a cero. Por tal razón, en el análisis del comportamiento global de las

estimaciones fue necesario calcular el valor absoluto del error porcentual (*AEP*), utilizando la siguiente ecuación:

$$AEP = \left| \frac{\text{Población estimada} - \text{Población censal}}{\text{Población censal}} \right| \times 100 \quad [11]$$

De este modo, el promedio de *AEP* resultó ser adecuado para evaluar la calidad de un conjunto de estimaciones y un buen parámetro de comparación de los resultados de dos o más conjuntos.

Por otra parte, se analizó el comportamiento de *AEP* con respecto a la magnitud de la población del cantón considerado. Para medir el efecto que tiene dicha magnitud, se calculó el promedio de *AEP* ponderado por la población censal:

$$\frac{\sum_j (AEP)_j \cdot P_j}{\sum_j P_j} \quad [12]$$

donde P_j es la población censal del cantón j .

Si el tamaño del cantón no afecta las estimaciones, dicho promedio no debió ser muy diferente del promedio simple de *AEP*.

7. Precisión de las estimaciones

De acuerdo con lo anterior, la precisión de las estimaciones se determinó mediante una comparación de su valor con la información censal. Para calificar una estimación como adecuada o inadecuada, se consideró que la información censal está sujeta a errores de cobertura. En el último censo, este error fue estimado aproximadamente en un 6% (CELADE/DGEC/MIDEPLAN, 1988). En consecuencia, no parece lógico que se les exija una extrema precisión a las estimaciones poblacionales de áreas menores. Por esta razón, para efectos del presente trabajo se consideraron “adecuadas” aquellas estimaciones que presentaron errores menores al 5%, “aceptables” las que estuvieron entre 5% y 10% e “inadecuadas” las que superaron el 10%.

II. RESULTADOS

1. Precisión y aplicación simple de las variables sintomáticas

A continuación se analiza el comportamiento de cada una de las variables, consideradas en el estudio en el período 1973-1984. En una primera etapa, se seleccionaron 22 cantones de las provincias de San José y Heredia, con respecto a cada uno de los cuales se recopiló información correspondiente a las siete variables consideradas en los últimos dos años censales (1973 y 1984). Se determinó, tanto para la población como para cada variable, la razón de cambio con respecto a la proporción que representa cada cantón en el total del país, en el período 1973-1984.

En el cuadro 1 se presenta la matriz de correlación de estas razones de cambio. Los mayores coeficientes de correlación con la población corresponden a los indicadores “padrón electoral” y “nacimientos”, y son superiores a 0.80. Las defunciones y la matrícula escolar también presentaron un coeficiente de correlación superior a 0.60, por lo que es importante considerar su aporte al proceso de estimación. Las otras tres variables presentan un coeficiente de correlación muy bajo, por lo que se decidió excluirlas del estudio.

Las variables sintomáticas seleccionadas se utilizaron en la segunda etapa con el objeto de hacer estimaciones poblacionales para todos los cantones del país.

Como ya se ha mencionado, para aplicar los métodos de correlación de razón y correlación de tasa se requiere información de los dos censos anteriores. El cálculo de las estimaciones poblacionales para 1984 exige utilizar los censos de 1963 y 1973, lo que obligó a trabajar con los 68 cantones existentes en el país en 1963. Por esta razón, la información correspondiente a los cantones creados después de este año se incluyó en el lugar de procedencia de cada uno. Este problema no afecta a los métodos de razón censal y diferencia de tasas, que son los únicos para los cuales se requiere información del censo de 1973, cuando había 79 cantones. Sin embargo, para efectos comparativos, se decidió considerar 68 en todos los casos.

Por otra parte, para dar mayor estabilidad a la estimación de los nacimientos y defunciones, se decidió calcular un promedio trienal, en lugar del valor simple, lo que reduce la variabilidad de estas variables sintomáticas.

Cuadro 1
MATRIZ DE CORRELACIÓN DE LAS RAZONES DE CAMBIO -
EN LAS PROPORCIONES DEL TAMAÑO POBLACIONAL Y DE OCHO
VARIABLES SINTOMÁTICAS EN 22 CANTONES DE SAN JOSÉ Y HEREDIA
1973-1984

	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
Población	1.00							
Padrón electoral	0.88	1.00						
Nacimientos	0.87	0.79	1.00					
Defunciones	0.65	0.58	0.56	1.00				
Matrícula escolar	0.61	0.36	0.48	0.21	1.00			
Asistencia centro de salud	0.35	0.38	0.10	0.47	0.12	1.00		
Permisos de construcción	0.23	0.07	0.00	0.04	-0.02	0.30	1.00	
Afiliación a la CCSS	-0.31	-0.18	-0.27	-0.23	-0.48	-0.07	0.09	1.00

Fuente: Elaboración propia sobre la base de información proporcionada por la Dirección General de Estadística y Censos (DGEC), Caja Costarricense de Seguro Social (CCSS), Ministerio de Educación Pública, Ministerio de Salud, Tribunal Supremo de Elecciones.

(a): Población; (b): padrón electoral; (c): nacimientos; (d): defunciones; (e): matrícula escolar; (f): asistencia a centro de salud; (g): permisos de construcción; (h): afiliación a la CCSS.

En el cuadro 2 se presentan los errores producidos en las estimaciones, para las cuatro variables sintomáticas consideradas.

Al igual que en la primera etapa, el padrón electoral, los nacimientos y el crecimiento vegetativo son las variables que dan los mejores resultados a este nivel. El padrón electoral es evidentemente superior a los demás indicadores. Al utilizar el padrón electoral en las estimaciones poblacionales, el error de estimación no superó el 5%, y ninguna de las demás variables sintomáticas se aproximó a este valor. Los nacimientos y el crecimiento vegetativo también ofrecen buenos resultados. Las estimaciones derivadas de la matrícula escolar y de las defunciones contienen grandes errores, en una proporción muy alta errores porcentuales que superan el 10%, por lo que estas estimaciones deben utilizarse con gran cautela.

Cuadro 2
**COMPARACIÓN DEL VALOR ABSOLUTO DEL ERROR PORCENTUAL
 DE (AEP) LAS ESTIMACIONES POBLACIONALES A NIVEL
 CANTONAL POR VARIABLE SINTOMÁTICA
 SEGÚN MÉTODO UTILIZADO, 1984**

Método	Variable sintomática			
	Padrón electoral	Naci- mientos	Matrícula escolar	Defun- ciones
Razón censal				
Promedio de AEP	5.0	8.8	11.3	15.9
Porcentaje de AEP inferior a 5%	61.8	42.6	30.9	14.7
Porcentaje de AEP superior a 10%	13.2	33.8	51.5	67.6
Diferencia de tasas				
Promedio de AEP	4.3	8.3	13.1	17.9
Porcentaje de AEP inferior a 5%	66.2	44.1	22.1	16.2
Porcentaje de AEP superior a 10%	5.9	32.4	58.8	72.1
Razón correlación				
Promedio de AEP	5.0	6.7	9.5	10.5
Porcentaje de AEP inferior a 5%	58.8	51.5	32.4	29.4
Porcentaje de AEP superior a 10%	13.2	20.6	41.2	36.8
Tasa correlación				
Promedio de AEP	4.9	6.4	9.6	10.4
Porcentaje de AEP inferior a 5%	55.9	51.5	36.8	27.9
Porcentaje de AEP superior a 10%	11.8	23.5	42.6	44.1

Fuente: Elaboración propia.

2. Aplicación de métodos de estimación con múltiples variables

A continuación se presentan los resultados obtenidos con más de una variable sintomática. En los casos de la razón censal y la diferencia de tasas, las estimaciones basadas en la combinación de dos o más variables corresponden al promedio de estimaciones simples. En los de correlación de razón y correlación de tasa, basta con incluir las variables pertinentes en el modelo de regresión múltiple.

En el cuadro 3 se indica la combinación de las variables que mostraron la mayor precisión en las estimaciones.

Cuando se utiliza el padrón electoral en combinación con los nacimientos como variables sintomáticas se obtienen buenos resultados, fundamentalmente cuando se emplean métodos basados en modelos de regresión. De este modo, el 72% de los cantones presentan un *AEP* inferior al 5%, un promedio de *AEP* de aproximadamente un 4% y una desviación estándar de los *EP* cercana al 5%. La combinación de la matrícula escolar, el padrón electoral y los nacimientos, reduce el promedio de *AEP* en 0.3 puntos porcentuales; en el caso de la correlación de tasa, disminuye el número de cantones con *AEP* superior al 10%.

Los resultados revelan cierta influencia de la magnitud de la población del área en las estimaciones, pues en todos los modelos se reduce el error promedio cuando se pondera por esta variable.

En el cuadro 4 se presentan los valores de R^2 correspondientes a los cuatro modelos del cuadro 3 y la significación de los coeficientes de cada variable sintomática, calculados mediante la prueba estadística “*t*”.

Todos los modelos empleados muestran un R^2 alto y los indicadores utilizados son significativos al nivel del 5%, como lo muestra la prueba “*t*”, con la excepción de la matrícula escolar. Este resultado señala que el aporte de la matrícula escolar al modelo que incluye el padrón electoral y los nacimientos es mínimo, dado que R^2 únicamente aumenta en un punto porcentual.

Con respecto al cumplimiento de los supuestos del modelo de regresión, las pruebas hechas demuestran que el modelo de correlación de tasa ofrece una mayor consistencia, debido fundamentalmente a que el método de correlación de razón presenta problemas con el supuesto de variancia constante (Chaves, 1997).

Cuadro 3

**COMPARACIÓN DEL VALOR ABSOLUTO DEL ERROR PORCENTUAL (AEP)
Y DEL ERROR PORCENTUAL (EP) EN LAS ESTIMACIONES
POBLACIONALES A NIVEL CANTONAL POR VARIABLE
SINTOMÁTICA SEGÚN MÉTODO UTILIZADO, 1984**

Variable sintomática utilizada	Método			
	Razón censal	Diferencia de tasas	Correlación de razón	Corre- lación de tasa
Padrón electoral				
Promedio de AEP	5.0	4.3	5.0	4.9
Desviación estándar de los EP	6.5	5.5	6.4	6.3
Promedio ponderado de AEP	4.8	4.3	4.6	4.6
Valor máximo de AEP	21.2	16.6	18.5	19.4
Porcentaje de AEP inferior a 5%	61.8	66.2	58.8	55.9
Porcentaje de AEP superior a 10%	11.8	5.9	13.2	11.8
Padrón electoral y nacimientos				
Promedio de AEP	4.7	4.7	4.1	4.0
Desviación estándar de los EP	6.8	6.1	4.9	4.8
Promedio ponderado de AEP	3.8	4.0	3.9	3.9
Valor máximo de AEP	21.6	20.7	12.1	12.5
Porcentaje de AEP inferior a 5%	66.2	64.7	72.1	72.1
Porcentaje de AEP superior a 10%	10.3	10.3	2.9	2.9
Padrón electoral, nacimien- tos y matrícula escolar				
Promedio de AEP	4.8	5.4	3.8	3.8
Desviación estándar de los EP	5.9	6.6	4.7	4.6
Promedio ponderado de AEP	4.1	4.8	3.7	3.6
Valor máximo de AEP	17.5	17.2	13.5	14.2
Porcentaje de AEP inferior a 5%	55.9	57.4	72.1	72.1
Porcentaje de AEP superior a 10%	8.8	14.7	2.9	1.5

Fuente: Elaboración propia.

Cuadro 4
VALOR DE R² Y SIGNIFICACIÓN DADA POR LA PRUEBA "T"
PARA LOS TRES MODELOS DE REGRESIÓN MÚLTIPLE
POR VARIABLE SINTOMÁTICA UTILIZADA, SEGÚN
MÉTODO DE ESTIMACIÓN
(NIVEL CANTONAL) ^a

Variables sintomáticas incluidas	Correlación de razón		Correlación de tasa	
	R ²	Prueba "t"	R ²	Prueba "t"
Modelo 1	0.76		0.74	
Padrón electoral		0.00		0.00
Modelo 2	0.81		0.81	
Padrón electoral		0.00		0.00
Nacimientos		0.00		0.00
Modelo 3	0.82		0.82	
Padrón electoral		0.00		0.00
Nacimientos		0.00		0.00
Matrícula escolar		0.07		0.06

Fuente: Elaboración propia.

^a Véanse los modelos de regresión presentados en el cuadro 3.

3. Comparación con estimaciones obtenidas mediante el método de crecimiento vegetativo

Como se ha mencionado, existe una gran demanda de estimaciones poblacionales basadas en el crecimiento vegetativo de la población, específicamente de las producidas por la Dirección General de Estadística y Censos de Costa Rica. Por lo tanto, se decidió comparar las estimaciones de la DGEC en 1984 con los resultados obtenidos mediante la aplicación de las variables sintomáticas. En el cuadro 5 se presenta un resumen de los errores observados al comparar dichas estimaciones con la información censal correspondiente a los diferentes cantones.

Cuadro 5
**COMPARACIÓN DEL VALOR ABSOLUTO DEL ERROR PORCENTUAL
 (AEP) DE LAS ESTIMACIONES POBLACIONALES A NIVEL
 CANTONAL, CON CRECIMIENTO VEGETATIVO
 1984**

	Estimaciones bajo un supuesto de crecimiento natural	Estimaciones con variables sintomáticas ^a
Promedio de AEP	12.4	4.0
Valor máximo de AEP	60.7	12.5
Porcentaje de AEP inferior a 5%	28.4	72.1
Porcentaje de AEP superior a 10%	44.4	2.9

Fuente: Elaboración propia.

^a Los resultados presentados corresponden a la aplicación del método de correlación de tasa con el padrón electoral y los nacimientos.

Estos resultados evidencian grandes diferencias en términos de precisión; las más precisas son las estimaciones hechas sobre la base de variables sintomáticas. Si se analiza únicamente el promedio de *AEP*, se observa un incremento de más de ocho puntos porcentuales. Esto señala que, al menos en el caso de períodos bastante alejados del año censal, las estimaciones calculadas a partir de variables sintomáticas ofrecen una mejor alternativa que simplemente suponer un crecimiento vegetativo.

CONCLUSIONES

Los resultados del estudio señalan que son pocas las variables sintomáticas confiables de que se dispone. A pesar de ello, es posible realizar estimaciones poblacionales a nivel cantonal de una calidad aceptable.

El número de electores inscritos en el padrón electoral y los nacimientos son las variables que presentaron la más alta correlación (cercana al 0.90) con respecto a la población. Además, éstas fueron las

que ofrecieron las mejores estimaciones, especialmente con las técnicas en las que se utiliza regresión lineal. Si se usan estas variables como predictoras, sólo 3% de los cantones presenta una desviación del valor censal superior al 10%, con un error promedio cercano al 4%.

En el presente caso las estimaciones se efectuaron once años después del censo. Para períodos poscensales más cortos cabría esperar estimaciones más exactas, debido a que las proporciones y las tasas de las variables sintomáticas no presentarían cambios tan acentuados.

Las diferencias entre los cuatro métodos estudiados no son muy marcadas, por lo que se podría utilizar cualquiera de ellos sin que se produjeran mayores variaciones en cuanto a la calidad de las estimaciones; sin embargo, las técnicas de correlación de razón y correlación de tasa son evidentemente más precisas.

Un elemento importante que queda en evidencia en el estudio, y que merece ser analizado más en detalle, es la distribución de los errores en las estimaciones de acuerdo con el tamaño poblacional. Los modelos de estimación demuestran que, mientras más reducida es la magnitud de la población, menor es la precisión de las estimaciones. Este hecho se observa claramente en la disminución del promedio de *AEP* cuando se pondera por el tamaño de la población de cada área menor, así como en las diferencias, ya comentadas, entre provincias, cantones y distritos. Este fenómeno puede deberse a muchas causas. En primer lugar podría atribuirse a la calidad de la información sintomática, que se va deteriorando a niveles desagregados menores (Rincón, 1989). Otra causa podría ser el hecho de que los errores dependen de la calidad de los cómputos censales. Sin embargo, la información censal está sujeta a problemas de cobertura, que difieren en magnitud de un área menor a otra. Lógicamente, es de esperar que esas diferencias sean más marcadas en unidades político-administrativas más pequeñas y en poblaciones más reducidas (Wolter y Causey, 1991). Una tercera causa que podría explicar las diferencias en lo que respecta a los errores —y que constituyó un serio problema en el manejo de información—, es la variación de los límites geográficos producida por la creación constante de nuevas áreas. Esta inestabilidad obliga a llevar a cabo reagrupaciones o desagregaciones de las localidades pertinentes, lo que puede provocar errores en el manejo de los datos sobre población y de la información sintomática.

Una de las metas implícitas del estudio era el desarrollo de una metodología que permitiera realizar estimaciones poblacionales de mejor calidad que la obtenida bajo el supuesto de un crecimiento

vegetativo de la población. Al comparar las estimaciones resultantes de la aplicación de los cuatro métodos con las publicadas por la Dirección General de Estadística y Censos sobre 1984, se observaron grandes diferencias en términos de precisión. El promedio de *AEP* fue de 12.4%, porcentaje que supera en alrededor de 8.5 puntos porcentuales las estimaciones hechas en el estudio.

Pese a lo anterior, y debido a la falta de información censal, la prueba de precisión de las estimaciones poblacionales tuvo que realizarse considerando el año 1984. Esto explica la necesidad de repetir el estudio tan pronto como se realice un nuevo censo de población, a fin de observar la estabilidad de las variables sintomáticas y de los métodos de estimación.

El presente estudio permite concluir que la exactitud de un método no puede ser generalizada y que, aunque un determinado procedimiento de estimación sea adecuado, no asegura que su aplicación a diferentes situaciones conduzca a resultados que reflejen fielmente la realidad. Al analizar las estimaciones se deben considerar las hipótesis sobre los métodos y la calidad de la información que se maneja. Es importante, entonces, tener un juicio crítico, y no caer en el uso mecánico de los métodos desarrollados ni de los resultados obtenidos.

BIBLIOGRAFÍA

- CELADE/DGEC/MIDEPLAN (Centro Latinoamericano de Demografía/ Dirección General de Estadística y Censos/Ministerio de Planificación Nacional y Política Económica) (1988), *Proyecciones Nacionales de Población, 1975-2025*, San José, Costa Rica.
- Chaves, Edwin (1997), "Indicadores sintomáticos en las estimaciones poblacionales para áreas menores. Costa Rica", tesis para optar al grado de Magister Scientiae, Universidad de Costa Rica, San José, Costa Rica.
- Ericksen, Eugene P. (1974), "A regression method for estimating population changes of local areas", *Journal of the American Statistical Association*, vol. 69.
- (1973), "A method for combining sample survey data and symptomatic indicators to obtain population estimates for local areas", *Demography*, N° 10.
- Feeney, Donald, James Hibbs y Thomas Gillaspay (1995), "Ratio-correlation method", *Basic Methods for Preparing Small-Area Population Estimate*, Norfleet Rives y otros (comps.), Laboratorio de demografía aplicada, Departamento de Sociología Rural, Escuela de Agricultura y Ciencias Biológicas, Universidad de Wisconsin-Madison, Servicio de Extensión.
- Mandell, Marylou y Jeffrey Tayman (1982), "Measuring temporal stability in regression models of population estimation", *Demography*, vol. 19.
- Namboodiri, Krishnan (1972), "On the ratio-correlation and related method of subnational population estimation", *Demography*, vol. 9.
- O'Hare, W. (1980), "A note on the use of regression methods in population estimates", *Demography*, vol. 17.
- Rincón, Mesa, Manuel José (1989), *Sistema para elaborar proyecciones subnacionales, por sexo y grupos de edades, por el método de los componentes*, San José, Costa Rica, Centro Latinoamericano de Demografía (CELADE).
- Rives, Norfleet W. (1982), "Assessment of a survey approach", *Population Estimates: Methods for Small Area Analysis*, Everetts S. Lee y Harold F. Goldsmith (comps.), Beverly Hills, California, SAGE Publications.
- Serow, William y Norfleet Rives (1995), "Small area analysis: assessing the state of the art", *Basic Methods for Preparing Small-Area*

- Population Estimate*, Norfleet Rives y otros (comps.), Laboratorio de demografía aplicada, Departamento de Sociología Rural, Escuela de Agricultura y Ciencias Biológicas, Universidad de Wisconsin-Madison, Servicio de Extensión.
- Simpson, Stephen y otros (1996), "Updating small area population estimates in England and Wales", *Journal of the American Statistical Association*, vol. 159.
- Swanson, David (1980), "Improving accuracy in multiple regression estimates of population using principles from causal modelling", *Demography*, vol. 17.
- Swanson, David y Lucky Tedrow (1984), "Improving the measurement of temporal change in regression models used for County population estimates", *Demography*, vol. 21.
- Voss, Paul R. y otros (1995), "Censal ratio methods", *Basic Methods for Preparing Small-Area Population Estimate*, Norfleet Rives y otros (comps.), Laboratorio de demografía aplicada, Departamento de Sociología Rural, Escuela de Agricultura y Ciencias Biológicas, Universidad de Wisconsin-Madison, Servicio de Extensión.
- Wolter, Kirk y Beverley Causey (1991), "Evaluation of procedures for improving population estimates for small areas", *Journal of the American Statistical Association*, vol. 86.